



DTIC FILE COPY

AD-A199 024

Diagnosis and Remediation in the Context of Intelligent Tutoring Systems

D. Sleeman, A.E. Kelly, R. Martinak,
R.D. Ward, and J. Moore

Stanford University

for

Contracting Officer's Representative
Judith Orasanu

ARI Scientific Coordination Office, London
Milton S. Katz, Chief

Basic Research Laboratory
Michael Kaplan, Director

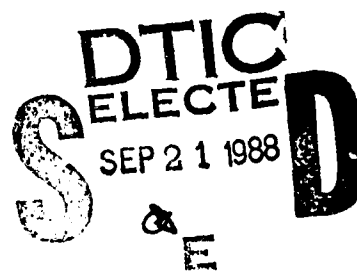


U. S. Army

Research Institute for the Behavioral and Social Sciences

July 1988

Approved for the public release; distribution unlimited.



U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the
Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

Stanford University

Technical review by

Steve Kronheim



Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

This report, as submitted by the contractor, has been cleared for release to Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or other reference services such as the National Technical Information Service (NTIS). The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS --	
2a. SECURITY CLASSIFICATION AUTHORITY --			3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.	
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE --				
4. PERFORMING ORGANIZATION REPORT NUMBER(S) --			5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 88-66	
6a. NAME OF PERFORMING ORGANIZATION Stanford University	6b. OFFICE SYMBOL (If applicable) --	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences		
6c. ADDRESS (City, State, and ZIP Code) Palo Alto, California		7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600		
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Same as 7.	8b. OFFICE SYMBOL (If applicable) --	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-84-K-0279		
8c. ADDRESS (City, State, and ZIP Code) --		10. SOURCE OF FUNDING NUMBERS		
		PROGRAM ELEMENT NO. 6.11.02.B	PROJECT NO. 2Q1611 02B74F	TASK NO. n/a
		WORK UNIT ACCESSION NO. n/a		
11. TITLE (Include Security Classification) DIAGNOSIS AND REMEDIATION IN THE CONTEXT OF INTELLIGENT TUTORING SYSTEMS				
12. PERSONAL AUTHOR(S) D. Sleeman, A.E. Kelly, R. Martinak, R.D. Ward, and J. Moore				
13a. TYPE OF REPORT Final Report	13b. TIME COVERED FROM 6/84 TO 9/87	14. DATE OF REPORT (Year, Month, Day) July 1988	15. PAGE COUNT 28	
16. SUPPLEMENTARY NOTATION Judith Orasanu, contracting officer's representative. Task performed under subcontract from Stanford University by University of Aberdeen, AB9 2UB, Scotland.				
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP		
			Intelligent Tutoring System, Reteaching	
			Artificial Intelligence, Cognition	
			Model-Based Remediation, PIXIE	
19. ABSTRACT (Continue on reverse if necessary and identify by block number)				
<p>This research note provides an overview of the four major aspects of the PIXIE project: the field work undertaken to determine how teachers diagnose and remediate during introductory algebra courses, the set of experiments run to determine the relative effectiveness of Model-Based Remediation (MBR) and Reteaching, systems work carried out to remedy shortcomings noted earlier in the Intelligent Tutoring System (PIXIE), and an experiment run to determine whether it is possible to enhance teachers' diagnostic capabilities.</p> <p>More detailed discussion of each of these topics are provided in four separate reports.</p>				
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Judith Orasanu			22b. TELEPHONE (Include Area Code) 202/274-8722	22c. OFFICE SYMBOL BR

ABSTRACT

This paper provides an overview of the 4 major aspects of the PIXIE project, namely: the field work undertaken to determine how teachers diagnose and remediate in introductory algebra; the set of experiments run to determine the relative effectiveness of Model-Based-Remediation (MBR) and Reteaching; systems work carried out to remedy shortcomings noted earlier in the Intelligent Tutoring System, PIXIE; and an experiment run to determine whether it is possible to enhance teachers' diagnostic capabilities. (More detailed discussions of each of these topics are provided in 4 separate technical reports).

The major conclusions from the four phases of the work are:

Field work: the teachers involved in the study, tutored algebra essentially procedurally.

Relative effectiveness of MBR and Reteaching: for algebra when taught procedurally with this age group, Reteaching seems as effective as MBR. This, in turn, implies that CAI is as effective as ICAI. Further we noted the importance of treating different types of errors differently; e.g., a consistent mal-rule should be treated differently to a slip.

System work: The initial basic PIXIE system has now been enhanced so that it can diagnose and remediate in several domains; use information of the student's intermediary working to reduce the number of remedial models presented to a student; and create a more global analysis of a student's performance.

Teachers as diagnosticians: this experiment concluded that exposure to the TPIXIE program did enhance the trainee teachers' ability to diagnose student errors.

The paper concludes with an extensive set of conclusions and suggestions for further work.

1. INTRODUCTION

Despite the considerable advances which have taken place in cognitive psychology, and in particular in information processing psychology, in the last two decades, the field does not have a prescriptive theory of instruction. Consequently, cognitive and instructional psychology are essentially still empirical sciences, although they have a growing corpus of knowledge to guide decisions. Several cognitive psychologists now view the field of intelligent tutoring systems (ITSs) as offering an important test bed for psychological theories (Anderson, et al, 1984); certainly these systems have the important characteristic of producing a reproducible environment. The lack of overall theory has led this research group to be particularly rigorous with field testing of its systems. This as we shall see has been a sobering exercise for the team, but, we hope, a valuable one for the field as a whole!

Given an accurate model of a student's performance in a domain (algebra), the focus of this project has been, how does one build an effective remedial system? The overall design assumed that remediation would be based on information in the student model, and that such a remedial system would be highly effective. It was then proposed to further fine-tune this remediation to tailor it to student's individual aptitudes, and learning styles. Indeed, we hoped to implement a truly adaptive intelligent tutoring system, namely one that would address the aptitude-treatment interaction issue (Cronbach & Snow, 1977). It was tacitly assumed that:

MODEL-BASED-REMEDICATION would be superior to RETEACHING.

In the early 1980's, due to the influence of the BUGGY work (Brown & Burton, 1978) and the carry over of the programming debuggy analogy, it

was generally accepted that:

- diagnosing a student's error was much more complex than (subsequent) remediation, i.e., remediation followed trivially once one had an accurate student model.
- highlighting a student's specific error(s) would create cognitive dissonance which would then make the student receptive to hearing the "truth".
- by and large it was expected that many student errors would be stable, i.e., students would have (reasonably) stable models of the task domain.

Brown and VanLehn (1980) suggest that the metaphor of the computer bug may have been misleading, and that bug migration is a phenomenon which the field needs to take seriously. Sleeman (1983) noted that there were different types of errors present in a population of algebra students, and that many students seem to follow a pattern of maturation during their understanding of a topic:

UNPREDICTABLE -> CONSISTENT USE of MAL-RULES -> CORRECT

This project has produced experimental evidence which challenges the assumptions listed above, and which supports the idea that students' errors vary over time and in duration.

Section 2 describes the studies undertaken to determine how teachers diagnose and remediate student errors in algebra; this section also includes a brief description of the remedial sub-system that was subsequently implemented. Section 3 describes a series of experiments undertaken to probe the effectiveness of the remedial sub-system; specifically, we investigated its effectiveness against simply reteaching.

Section 4 describes some modifications carried out to the PIXIE system to make it a more effective tutor. Section 5 describes experiments that measure attempts to enhance teachers' diagnostic capabilities. Section 6 reports the overall conclusions of the research, and section 7 sets out an ambitious program of work which follows from this study and its conclusions.

2. FIELD STUDIES OF TEACHERS CARRYING OUT DIAGNOSIS & REMEDIATION

In order to begin to identify what makes for effective diagnosis and remediation of linear algebraic equations, and how this relates to the design of intelligent tutoring systems, two substantial and two supportive studies of master teachers were undertaken. ^{In the first study,} ~~Partially~~ 4 experienced teachers were shown a series of task-answer pairs which had been incorrectly worked by pupils, and asked to suggest a diagnosis and a suitable remediation. Although there was often a common error in each of the several sets of tasks presented, this was not pointed out to the teachers. Only one of the four teachers looked for a common error; the others were happy to make suggestions on a task-by-task basis. The teachers suggested remediation for approximately 50% of the errors, it being notable that when multiple errors occurred the teachers only suggested remediation for one of them (the most important error?). Further, procedural forms of remediation were suggested more than twice as frequently as conceptually-based forms of remediation. For further details of this study see Kelly & Sleeman (1986).

In the second study an experienced maths teacher was observed tutoring eight students, based on the diagnosis provided for each student by the PIXIE system. This teacher's remediation was also essentially procedural but it did have two striking and unexpected features. Firstly, this teacher having been told that the student was doing flipped division (i.e., transforming tasks of the form $5x=3$ to $x=5/3$) would probe this diagnosis by means of a series of simpler equations to determine the reason for this. For instance, did the student know how to write 5 divided by 3? Did he know how to cope with improper fractions? Or was he simply lacking a general procedure to solve tasks of this form? Having carried out this further probing and diagnosis, the teacher would

then proceed to give the student procedurally-based remediation. Because of the way in which these diagnoses had been confirmed, we refer to this as causal-based-remediation. Secondly, the teacher presented his remediation in a very tentative way; taking great care to point out to the student the steps he had done correctly, and the reasonableness of the errors made. This teacher was a model empathetic tutor.

The first supporting study was a series of interviews with 3 Irish mathematics teachers; the interviews covered how they taught and remediated algebra bugs. All supported the need to teach algebra (and one suspects most of mathematics) procedurally on the grounds of effectiveness and time-constraints. Further, these teachers also stressed the need not "to demolish a student's confidence by pointing out a series of errors".

For the second supporting study, we held a workshop for algebra teachers in the San Francisco area to discuss the teaching and remediation of algebra and had all our earlier observations about the centrality of procedural teaching confirmed.

As a result of these studies we concluded that the vast majority of teachers taught algebra procedurally. It was therefore decided to implement a remedial system which had that form. (It was decided not to implement the more complex causal-based-remediation approach.)

The philosophy behind the remediation was that it would be beneficial to highlight each error, to explain what is wrong, and what should have been done. Finally, to reinforce these points the task was reworked correctly, with a commentary on each of the steps. Throughout this document, this will be referred to as model-based-remediation (MBR); an alternative name being error-based-remediation (figure 1). Another com-

monly used mode of remediation available within RPIXIE is Reteaching, i.e., the latter half of the protocol in figure 1 (namely, that below the line).

The figure illustrates interaction with the RPIXIE system during the series of experiments described in section 3. The system developments described in section 4 were carried out in parallel with this experimental work, and therefore had not been completed at the time the instructional experiments were conducted.

FIGURE 1
An Example of a Student Interaction: MBR

(This protocol is slightly edited)

Solve (f = Finished, q = Quit early)

$$7x = 2 * 2$$

S: $7x = 4$

S: f

Simplify further $7x = 4$

S: $x = 4$

S: f

Thank you,
but you didn't get that one right.
We will now look at that task again.

Here we are considering the task $7x = 2 * 2$

Did you work out the task something like this?:

$$7x = 2 * 2$$

$$7x = 4$$

$$7x/7 = 4$$

$$x = 4$$

Please type YES or NO: y (Student replies)

It seems that you changed $7x = 4$ to something like $7x/7 = 4$

This is WRONG.

Do NOT only divide ONE side by the coefficient.

You should divide both sides of the equation by the coefficient to change

$$7x = 4 \text{ to } 7x/7 = 4/7$$

Now let me show you how you SHOULD have worked the whole task

$$7x = 2 * 2 \quad \text{multiply 2 numbers}$$

$$7x = 4 \quad \text{divide both sides of the equation by the coefficient}$$

$$7x/7 = 4/7 \quad \text{divide}$$

$$x = 4/7 \quad \text{Finished}$$

3. THE SERIES OF EXPERIMENTS ON THE RELATIVE EFFECTIVENESS
OF THE SEVERAL REMEDIAL TREATMENTS

As noted in the introduction, the relative effectiveness of different forms of remediation was the central issue in this research. The intention was to build a highly adaptive, intelligent tutoring system. As a first step in this process, we attempted to verify the hypothesis that MBR (Model-based-remediation) was superior to Reteaching. Subsequent experimentation was to establish the optimum conditions for students with differing aptitudes.

Essentially, we could find no evidence supporting the greater effectiveness of MBR for algebra when taught procedurally (or more specifically, not for our target population). The rest of this section discusses in some detail the main points of the experiments conducted to investigate this issue. See Martinak, Sleeman, Kelly, Moore & Ward (1987) for a more detailed description of this series of experiments.

After a series of pilot studies to verify that students were able to easily use the RPIXIE system, we ran our first formal experiment. This, and the subsequent studies followed a pretest-intervention-posttest design. For a class of 24 13-14 year old pupils who were below average in mathematics, it was found that MBR and Reteaching by RPIXIE were both more effective than merely telling the student whether the task had been worked correctly. However, MBR was not better than reteaching; the performance of these groups were comparable. This was a surprising result.

This result led us to believe that the issues of remediation were much more subtle than initially suspected, and therefore we decided to replicate the study using human tutors. This second study gave essentially the same result. It was then hypothesised that these results may have

occurred because the treatments had not involved the students sufficiently in the remediation, or that alternatively, PIXIE's corrective comments, targetted at those part(s) of the task the student had worked incorrectly, were failing to create the expected cognitive dissonance. A third experiment was therefore conducted with 4 treatment groups, namely MBR, MBR + Cognitive Engagement (here the student was asked to reteach to the tutor the correct procedures), MBR + Cognitive Dissonance (here the student was required to substitute his (incorrect) solution back into the original equation, thereby demonstrating that his solution was wrong) and Reteaching. Again the results for all 4 groups were comparable.

This additional puzzling result led to a further range of hypotheses; specifically, to suppose that many errors are in fact unstable, that is, the same student given a comparable task on different occasions would work the task differently. Indeed, a retrospective analysis of the last experiment, showed that only 18-26% of errors made on the pre-test were present on the same items one week later during the tutorial. (Please note that this is a very stringent requirement for stability of errors; a more lenient criterion is introduced later.)

The fourth experiment in the series was explicitly designed to investigate the issue of stability. A test measure containing 51 items was developed - 17 sets of 3 comparable items. This measure was given twice at a week's interval. The intent of this study was to identify errors that were stable over time, and then to provide human tutoring on those. On this occasion, for an error to be classified as "stable", it had to occur at least twice on both pretests. Students with stable errors were assigned randomly to one of three conditions, namely MBR, Reteach or the control group. Both the MBR and Reteach groups were tutored individu-

ally for a 50 minute period; the control group took only the 2 pretests and the posttest. Below we give the average number of occurrences of the 19 most common errors for the 3 groups (these 19 errors account for 80% of the errors in this study):

	Pretest 1	Pretest 2	Posttest	Number of students in group
MBR	19.2	18.5	10.3	9
Reteaching	29.4	22.6	9.9	8
Control	32.0	26.4	26.0	8

These figures suggest that errors are fairly stable from Pretest-1 to Pretest-2, however, errors decrease substantially from the pretest to tutoring, presumably due to the effects of tutoring. An error once tutored, tends not to reappear in the same tutorial session; additionally, tutoring appears to suppress attentional errors*. These results also show that there are significantly fewer errors on the posttest for the treatment groups when compared with the control group; again both treatment groups were highly comparable. A further analysis of the data given in the above table shows that the percentage decrease in the number of stable errors between the first pretest and the posttest for MBR, Reteaching and the control group was respectively 46%, 66% and 19%. This suggests that although some errors are unstable, tutoring is effective at remediating stable errors, but again MBR is not more effective than reteaching. (These observations are consistent with Sleeman (1983) who reported an experiment in which the MBR group greatly out performed the control group.)

Several additional experiments were run with RPIXIE which generally supported the result that MBR and Reteaching were very comparable; see Mar-

* Errors caused by lack of attention to the task.

tinak, et al. (1987) for details.

These results will now be interpreted within the framework of the assumptions listed in the introductory section. Explicitly, the results from our experiment will be related to each of these assumptions.

Assumption 1. (Diagnosing a student's error is much more complex than remediation.) Even if a diagnosis has been made correctly, remediation involves conveying that information to the student in a way that is intelligible. Much of our social knowhow is about communication, e.g., phrasing a request so that it will appear attractive to the hearer etc. Remediation is no less subtle; the teachers in our study (section 2) seemed to understand that. (Unfortunately, RPIXIE did not!)

Conclusion: Those of us who have been enamoured with the technicalities of inferring student models, had overlooked the complexities inherent in subsequently communicating the remediation. (Note: this is not to say that diagnosis is a simple matter).

Assumption 2. (Highlighting a student's specific error(s) would create cognitive dissonance.) This set of experiments clearly established that, for this topic and teaching approach, reteaching and model-based remediation was better than no treatment at all, but that reteaching and model-based-remediation were highly comparable. This initially surprising result indicates that, for this topic and students, CAI would have been just as effective as ICAI (as, of course, CAI programs are quite capable of storing pre-worked solutions to tasks). Secondly, one interpretation of the fact that students did equally well on Reteaching as on MBR is that the students in the Reteaching group were self-correcting. That is, they compared their incorrect working with the correct form, and generally inferred their own errors. Again this

interpretation is consistent with other experiments on "passive" versus "active" instruction, and is consistent with the literature on meta-cognition, (Brown, 1978).

This explanation would explain why immediate feedback is so important for learning (Lewis & Anderson, 1985). (If the critical component is the provision of virtually instant feedback then this would explain why the feedback provided by teachers on exercises a week or so after the event is also not very effective.)

Assumption 3. (Many student errors would be stable.)

These experiments have supplied further evidence for the series for error-types suggested by Sleeman, (1983). That is, one should expect to find students with a range of types of errors, including:

- strongly held consistent mal-rules.
- related "families" of mal-rules which are applied "randomly".
- passing attentional errors (like adding/omitting signs).
- guesses because the tutor or the program demands an answer.*
- mental-slips and casual (typing) errors.

When the investigators reviewed their tapes with this classification in mind they found strong supporting evidence for it, and reported that it was clear that students had varying confidence concerning the correctness of the different types of errors. This analysis has considerable implications for remediation. Clearly, one might wish to highlight and

* After the 1981 experiment, a facility was added to PIXIE to allow students to QUIT any task, so as to avoid this situation.

discuss in detail a known stable error, but a detailed discussion of a pure guess might be counter-productive as it might help "cement" the incorrect form. How to phrase remedial comments, as we have seen, is also of vital importance. The version of RPIXIE used in these experiments lacks the sophistication of being able to make a "global" diagnosis of a student's error pattern. However, the analysis of these experiments suggests that this may be an important issue. Section 4 discusses a pilot system which produces more global diagnoses, i.e., diagnoses which "explain" a series of errors - possibly which occurred in various task-sets.

Further, the above analysis led to the suggestion that because the students had been taught procedurally they might not have acquired an (overall) mental model for the domain. We further hypothesized that had they been taught conceptually, then there would have been a greater chance of the student forming a mental model, and thus such students should exhibit more stable errors. We were unable to find any Aberdeen secondary schools that taught algebra conceptually. So this hypothesis remains untested.

The implications of the series of diagnostic/remedial experiments are discussed in some detail in sections 6 & 7.

4. SYSTEMS WORK

For the record, at the start of the project, the PIXIE system existed on a multi-user PDP10 system, and has subsequently been transferred to a variety of personal computers, the IBM XT, the Tektronix 4404, and finally to a SUN 3/52. The project has insisted, perhaps wrongly, that the system should remain in LISP. The IBM version was abandoned because the remedial system ran far too slowly under IQ-LISP (the promised compiler was not forthcoming). A combination of speed and technical problems with the Tektronix 4404 led us to transfer to the SUN system.

During the course of the 3-year project, an extensive amount of systems work has been carried out, Moore & Sleeman (1987). (Note these developments were completed after the experimental work described in Section 3). Below, work of principally educational importance is mentioned:

- the PIXIE shell has been modified so that it is possible to tutor (i.e., diagnose and remediate) in several subject areas. This gives the capability of having found a consistent precedence bug in algebra (e.g. $4+5x=19 \Rightarrow 9x=19$) to have the student tutored on arithmetic precedence i.e., tasks of the form $3+4*5$.
- The remedial system has been improved so that it selects remedial models which are consistent with the student's intermediary workings to present to the student. PIXIE had the ability to infer a set of models which are consistent with the student's answer. However, RPIXIE only proposed MBR if it had inferred only a single model. When it had multiple consistent models it simply retaught the task. Using the student's intermediary working the set of models can often be greatly reduced; this reduced subset is now presented to the student by the enhanced system.

- A sub-system has been implemented which produces a more global analysis of a student's performance on a wide range of tasks. Previously, the most commonly used mode of the RPIXIE system produced a diagnosis (and if needed remediation) which was specific to a particular task. This was too myopic a view. The current sub-system when it is shown a student analysis record of the following form:

$$5x=15 \Rightarrow x=3 \quad \text{and} \quad 5x=7 \Rightarrow x=2$$

suggests that it is probable the student can correctly solve tasks of the form $ax=b$ when b is divisible by a , but not when b is indivisible by a . This sub-system also suggests sets of tasks that should be used in tutoring such a student.

Similarly, given the following student performance:

$$5x+3=11 \Rightarrow x=8/5 \quad \text{and} \quad 5x+3x=11 \Rightarrow x+x=11-3-5$$

this subsystem would suggest that the student can successfully solve tasks of the form $ax+b=c$, but not those of the form $ax+bx=c$, suggesting that the student does not know how to combine x -terms.

Various software aides have been produced for the developer of new knowledge bases. These include a program which, given the template for a level and the set of models, generates the set of most discriminating tasks. (Ideally these tasks would be completely discriminatory.) Another package checks for syntax errors and certain semantic inconsistencies in knowledge bases (e.g., entities being referenced but not defined.)

Although, not sponsored by this project, we have implemented during this period a system INFER*, which is able to infer mal-rules from previously unknown protocols, given additional background knowledge and some focusing heuristics. Additionally, we have implemented a system, MALGEN, which applies perturbations to correct rules, and filters out "variants"

which violate certain meta-constraints. For details of these approaches see Sleeman (1982) and Sleeman, Hirsh & Kim (1987).

The critical issue of field-testing these new sub-systems, and the subsequent integration of these several components into a further enhanced PIXIE-system is discussed in section 7.

5. AIDES FOR HELPING TEACHERS BE BETTER DIAGNOSTICIANS

The TPIXIE program drew some of its inspiration from the BUGGY program (Brown & Burton, 1978) which presents trainee teachers with incorrectly worked subtraction tasks and then asked them to suggest additional tasks and indicate how that same student, if consistent, would work them. The major difference between the BUGGY and TPIXIE is the domain of application.

A pilot study with the system in California, showed that trainee-teachers who used TPIXIE were somewhat better than those in the control group who merely worked algebra tasks. However, the trainee-teachers suggested that the example-set be changed so that more difficult tasks would be encountered earlier in the session. Also the analysis of the data showed that the transfer of knowledge to new but highly analogous tasks was not very substantial (Schneider, Kelly, Blando, Martinak, Sleeman & Snow, 1986).

A further experiment with an enhanced TPIXIE system was conducted in Aberdeen with a larger sample of trainee-teachers; for details of the system and the study see Kelly, Sleeman, Ward & Martinak, 1987. The encouraging trend of the pilot study was confirmed. The subjects on TPIXIE were significantly better at diagnosing algebra errors on the posttest than those in the control group. The study also recommended further refinements to the methodology and test instrument prior to replication.

If, as section 3 suggests, Reteaching is as effective as MBR, then there is less point in training teachers to be good diagnosticians than we had previously thought. Nevertheless, one could make the case, that being aware of possible student errors would make them better classroom teach-

ers; the implications of the TPIXIE project are further discussed in sections 6 and 7.

6. CONCLUSIONS

Listed below are the conclusions drawn from a series of PIXIE related studies:

- Virtually all teachers encountered in this study in American, English, Irish and Scottish schools taught algebra procedurally. (Section 2)
- Model-based-remediation and reteaching using humans as tutors are both more effective than no tutoring. (Section 3)
- Model-based-remediation and simply reteaching are equally effective when the tutoring is carried out by humans. This leads to the hypothesis discussed in section 3 that the students in the reteaching group were self-correcting, and the conclusion that, for some domains and some student populations, CAI would be as effective as ICAI. (Section 3)
- In the last study, a significant number of students had stable errors which accounted for approximately 80% of errors recorded. There appeared to be a bigger percentage of unstable errors when students interacted with the computer, namely with RPIXIE (section 3).
- There is further evidence that students make a wide variety of types of errors (from "hard" bugs to careless (typing) errors) and that students hold beliefs of varying strengths about these error types. (Section 3; see paragraph on Assumption 3).
- The PIXIE system has been further enhanced, so that it should be

more human-like in its tutoring - having the ability to tutor in several domains and to form "global" diagnoses. (These facilities now need to be thoroughly field-tested.) (Section 4)

- It is possible to train teachers to diagnose error patterns in examples wrongly worked by students. (Section 5)

7. FURTHER WORK SUGGESTED BY THIS STUDY DIAGNOSIS AND REMEDIATION

- An extensive set of field-trials is required to determine under what conditions Reteaching is as effective as model-based-remediation. Educators, as well as those in the ITS field, need to know how this is influenced by subject domain, age of student and teaching approach. (Probe whether conceptual teaching leads to more stable mental models).
- Run a study to further investigate the effectiveness of an MBR + Cognitive Dissonance condition, modelled after Swan (1983).
- Replicate the study to investigate the stability of errors in algebra with a larger N and with the requirement that each stable error should be represented in all conditions.
- Run a study to compare rates of attentional errors with human and computer tutoring (both MBR and Reteaching).
- Investigate how the stability of errors and models might be influenced by subject-domain, student age, level of attainment and teaching approach. As a secondary issue, one would wish to investigate the extent to which students have a distinguishable conceptual model and whether the range of error-types found in algebra are present in other domains.
- Run an experiment in which the student is distracted immediately after he has done a task, and before he is shown the Reteaching. It was suggested above that one reason why reteaching was as effective as model-based-remediation, might be because the student was essentially self-correcting. [If this hypothesis is correct the "distracted" stu-

dents would do considerably worse than those who are not.] Alternatively, run a study in which there is a differential time gap between working the task and receiving feedback.

- Get tutors to review the extensive working from a student and articulate a "global" diagnosis; have the tutor remediate a student on the basis of this analysis. Compare the effectiveness of this remediation with Reteaching.

- Compare (human) empathetic tutoring with "neutral" tutoring, ensuring that the instructional context of the material tutored is identical. [This experiment would need to be run for a variety of personality types as well as for the factors noted earlier.]

- Compare the effect of a human tutor giving detailed causal-based-remediation (see definition in section 2) against "straight" reteaching.

System

- Run extensive field trials to determine the effectiveness of the multi-domain diagnosis/remedial system, and of the system which can form global diagnoses. [IF this is successful, then a system should be implemented which integrates the higher-order diagnoses, multiple knowledge bases, as well as the INFER* algorithm (which is able to infer previously unknown mal-rules from protocols).]

TPIXIE (Studies to see if teachers can be taught to diagnose)

- Repeat the TPIXIE current study with a refined instrument; and

investigate again whether transfer is effective.

8. POSTSCRIPT

As a result of this study, some clear questions have evolved, which should be answered before it is sensible to build an Intelligent Tutoring System, namely:

- can human tutors demonstrate that MBR is more effective than Reteaching in that domain?
- are student errors in the proposed domain stable?

REFERENCES

Anderson, J. R., Boyle, C. F., Farrell, R. & Reiser, B. (1984). Cognitive Principles in the design of Computer Tutors. Proc. of the 6th Annual Conference of the Cognitive Science Program, Carnegie-Mellon University, Pittsburgh, PA. pp 2-9.

Brown, A. L. (1978). Knowing when, where and how to remember: A problem of metacognition. In R. Glaser (Ed.), Advances in Instructional Psychology, Volume 1. Hillsdale, N. J.: Erlbaum. pp77-165.

Brown, J. S. & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 2, pp 155-192.

Brown, J. S. & VanLehn, K. (1980). Repair Theory: a generative theory of bugs in procedural skills. Cognitive Science, 4. pp 379-426.

Cronbach, L. J. & Snow, R. E. (1977). Aptitudes and Instructional Methods. New York: Irvington

Kelly, A. E. & Sleeman, D. (1986). A study of Diagnostic and Remedial Techniques used by Master Algebra Teachers. Technical Report AUCS/TR8708, Department of Computing Science, University of Aberdeen.

Kelly, A. E., Sleeman, D., Ward, R. D. & Martinak, R. (1987). TPIXIE: A computer program to teach diagnosis of algebra errors. Technical report AUCS/TR8710, Department of Computing Science, University of Aberdeen.

Lewis, M. W. & Anderson, J. R. (1985). Discrimination of operator sche-

mata in problem solving: Learning from examples. Cognitive Psychology, 17, pp 26-65

Martinak, R., Sleeman, D., Kelly, A. E., Moore, J. & Ward, R. D. (1987). Studies of Diagnosis & Remediation with High School Algebra Students. Technical Report AUCS/TR8711, Department of Computing Science, University of Aberdeen.

Moore, J. & Sleeman, D. (1987). Enhancing PIXIE's tutoring capabilities. Technical Report AUCS/TR8709, Department of Computing Science, University of Aberdeen.

Schneider, B., Kelly, A. E., Blando, J. A., Martinak, R., Sleeman, D. & Snow, R. E. (1986). TPIXIE: Towards improved diagnosis of algebra error patterns by teachers. Proceedings of annual meeting of the American Psychological Association, Washington.

Sleeman, D. (1982). Inferring (mal)rules from pupil's protocols. In Proceedings of the 1982 European AI Conference, pp 160-164. (Republished in Proceedings of the International Machine Learning Workshop, Illinois, June 1983).

Sleeman, D. (1983). Basic algebra revisited: a study with 14 year olds. HPP technical report 83-9, Computer Science Dept, Stanford. (And republished in International Journal of Man-Machine Studies, (1985) pp 127-149.)

Sleeman, D., Hirsh, H. B. & Kim, In-Yung. (1987). Expanding an incomplete domain theory: Two-Case Studies. Technical Report, AUCS/8704.

Department of Computing Science, University of Aberdeen.

Swan, M. B. (1983). Teaching decimal place value. A comparative study of conflict and positive only approaches. Research Report 31, Shell Centre for Mathematical Education, University of Nottingham.